(17:32:19) dr ttol: you've imported all the ones i gave you
(17:32:33) David Gucwa: yes
(17:32:37) David Gucwa: in the process of, at least
(17:32:41) dr ttol: ok
(17:32:43) dr ttol: yee@bu.edu
(17:32:45) dr ttol: ab1036
(17:34:00) David Gucwa: ok I started that one
(17:34:22) dr ttol: check to make sure the logins are valid
(17:34:23) dr ttol: im not sure
(17:34:35) dr ttol: barroseu@bc.edu
(17:34:39) dr ttol: eutychius
(17:34:50) dr ttol: beelzebub@brown.edu
(17:34:53) dr ttol: asdfgh
(17:36:50) dr ttol: hoganc@alum.mit.edu
(17:36:52) dr ttol: csfb
(17:36:58) dr ttol: porcella@princeton.edu
(17:36:58) dr ttol: angelo
(17:37:31) dr ttol: carlsson@stanford.edu
(17:37:35) dr ttol: fake_pwd
(17:37:46) dr ttol: maria.rocha_oliveira@tufts.edu
(17:37:47) dr ttol: maria
(17:37:54) dr ttol: Stinab8@ucla.edu
(17:37:55) dr ttol: divya
(17:38:15) dr ttol: Sen4@georgetown.edu
(17:38:20) dr ttol: sexyrena
(17:38:34) dr ttol: rl6@duke.edu
(17:38:35) dr ttol: nacho
(17:38:38) dr ttol: some may not work
(17:41:04) David Gucwa: k
(17:44:45) David Gucwa: yeah 3 or 4 didn't work
(17:44:47) David Gucwa: I started the rest
(17:45:05) dr ttol: which ones didnt
(17:46:39) David Gucwa: rl6@duke.edu Sen4@georgetown.edu porcella@princeton.edu hoganc@alum.mit.edu
(17:47:59) David Gucwa: it's got 46,000 id's so far and still going
(17:48:08) David Gucwa: in total from all of them
(17:48:48) dr ttol: we need the profiles
(17:48:52) dr ttol: not just ids
(17:49:45) David Gucwa: I know
(17:50:10) dr ttol: singer.d@neu.edu
(17:50:12) dr ttol: 1211
(17:50:39) dr ttol: mj405@nyu.edu
(17:50:41) dr ttol: rajubros
(17:52:09) dr ttol: how many proiles?
(17:52:21) David Gucwa: none yet, I'll start that
(17:52:39) dr ttol: ok
(17:55:33) dr ttol: let me know when we hit 100 profiles
(17:56:54) David Gucwa: we've probably hit that by now
(17:57:19) dr ttol: we're prob setting up huge alarms at thefacebook
(17:57:30) dr ttol: how can we go faster
(17:58:22) dr ttol: procella@princeton.edu
(17:58:23) dr ttol: frankmaria
(17:58:56) David Gucwa: I'm not sure we can go any faster
(17:59:14) dr ttol: what if we had the scripts running from other computers at te same time
(17:59:32) David Gucwa: do we have the proxies available yet
(17:59:36) dr ttol: yes

(17:59:54) dr ttol: unix15.dmbhosting.com
(17:59:58) dr ttol: i2hub / Importer123
(18:00:05) dr ttol: 66.96.217.229
(18:00:13) dr ttol: i2hub / TX6ADGXIAHR
(18:00:19) dr ttol: 67.131.250.102
(18:00:23) dr ttol: i2hub / 4it2AXqi
(18:00:28) dr ttol: 67.18.33.226
(18:00:29) David Gucwa: do these have ssh access
(18:00:33) David Gucwa: I can't seem to get into the first one
(18:00:34) dr ttol: i2hubtes / Importer123
(18:00:41) dr ttol: 69.56.226.102
(18:00:44) dr ttol: i2hub / importer12
(18:00:51) dr ttol: i dont know about ssh
(18:00:52) dr ttol: some may
(18:00:53) dr ttol: some may not
(18:00:55) dr ttol: all have ftp
(18:01:04) David Gucwa: hm
(18:01:09) David Gucwa: ok
(18:07:07) David Gucwa: procella@princeton.edu doesn't work
(18:10:29) dr ttol: porcella
(18:10:32) dr ttol: cameron typo'd
(18:10:37) David Gucwa: ok
(18:12:32) dr ttol: cmg25@georgetown.edu
(18:12:35) dr ttol: top59gun
(18:16:47) dr ttol: how many profiles so far?
(18:19:07) David Gucwa: 5300
(18:19:27) dr ttol: what schools?
(18:20:00) dr ttol: or is that combined
(18:20:46) David Gucwa: it was all from one school, but I'm checking them now and the emails aren't there
(18:20:55) David Gucwa: they must have just gotten turned off
(18:21:04) dr ttol: fuck
(18:21:30) dr ttol: what school
(18:21:52) David Gucwa: umich
(18:22:16) dr ttol: they just got turned off?
(18:23:25) David Gucwa: wait not umich
(18:23:36) David Gucwa: yale
(18:23:58) David Gucwa: maybe they were never turned on, I'm not sure I checked
(18:24:01) David Gucwa: I thought I did though
(18:24:02) dr ttol: ok
(18:24:13) dr ttol: try the ones with emails
(18:24:16) dr ttol: not the ones without
(18:25:22) David Gucwa: ok I'm going to try bu
(18:25:26) David Gucwa: I just made sure the emails are showing
(18:25:26) dr ttol: ok
(18:26:25) dr ttol: see how many we can get
(18:26:37) dr ttol: did you try the other shells
(18:27:16) David Gucwa: yeah I've got one importing from princeton
(18:27:22) David Gucwa: most of them didn't have ssh open
(18:27:30) dr ttol: you need ssh?
(18:27:35) dr ttol: you said ftp is ok
(18:28:13) David Gucwa: yeah ftp will work
(18:28:19) David Gucwa: I just need to modify some things
(18:28:31) David Gucwa: ssh is just easier
(18:35:36) dr ttol: status
(18:35:38) dr ttol: on bu caching
(18:37:00) dr ttol: do the pton and gtown emails first

(18:37:05) dr ttol: the passwords will change soon
(18:37:48) dr ttol: just make sure they are cached tonight or something
(18:37:50) dr ttol: doesnt have to be now
(18:37:52) dr ttol: bu/umass are priority
(18:41:31) dr ttol: sk1281@columbia.edu
(18:41:32) dr ttol: khalid1
(18:42:55) David Gucwa: it's really hard to get a progress indication now because everything on the server is going pretty slowly
(18:43:07) dr ttol: which server
(18:43:07) dr ttol: theirs?
(18:43:10) David Gucwa: ours
(18:43:18) dr ttol: connectu?
(18:43:21) David Gucwa: i2hub
(18:43:27) dr ttol: where are you storing the sql
(18:43:30) David Gucwa: www.i2hub.com seems to load just fine in my browser though
(18:43:34) David Gucwa: no sql yet
(18:43:39) dr ttol: oh
(18:43:41) David Gucwa: I'm just dumping everything to files
(18:43:42) dr ttol: where are you caching the profiles
(18:43:43) dr ttol: oh
(18:43:45) dr ttol: why
(18:44:09) David Gucwa: makes it a lot easier to move them around from server to server
(18:44:17) David Gucwa: if we want to distribute it
(18:44:19) dr ttol: how are you going to import it
(18:44:36) David Gucwa: just read from the file
(18:44:51) dr ttol: you dont think relational db's are faster than file writing/reading?
(18:46:02) David Gucwa: I can put them into a db eventually if we decide that's faster. I just wanted to have them stored right now in the easiest possible way to work with.
(18:46:45) dr ttol: ok
(18:47:08) dr ttol: the writing to hard drive might be the reason its slow
(18:48:06) David Gucwa: Hm. Probably.
(18:48:28) David Gucwa: it's either IO or bandwidth
(18:48:32) David Gucwa: hard to tell which
(18:49:04) David Gucwa: we've imported about 10,000 profiles from bu
(18:49:24) dr ttol: with emailos?
(18:49:43) David Gucwa: they shut off
(18:49:55) dr ttol: ?
(18:50:03) dr ttol: what shut off
(18:50:03) David Gucwa: I think they have something that checks for this kind of crawling and turns off the emails
(18:50:55) dr ttol: how many emails did we get
(18:52:54) David Gucwa: checking
(18:53:15) David Gucwa: 341
(18:53:23) David Gucwa: That's pretty terrible
(18:53:35) dr ttol: yeah
(18:54:43) dr ttol: we're going into a conf call
(18:54:46) David Gucwa: ok

Conversation with drttol at 2005-02-20 09:08:31 on David Gucwa (aim)
(09:08:42) David Gucwa: I got your page
(09:09:34) dr ttol: hi
(09:09:45) dr ttol: i tried looking everywhere for your pager email
(09:09:48) dr ttol: where is it in the code?
(09:09:55) David Gucwa: in one of the dbs
(09:10:00) dr ttol: damn it
(09:10:09) dr ttol: i was going through the entire code
(09:10:17) David Gucwa: what's wrong with the importer?
(09:10:22) dr ttol: check this
(09:10:30) dr ttol: http://www.i2hub.com/dgucwa/face/grabFacebook.php?username=rdegutis@student.umass.edu&password=i2hub123
(09:11:11) dr ttol: ERROR:Thefacebook:Couldn't fetch user ID from main page.
(09:13:00) David Gucwa: it's taking a long time
(09:13:08) dr ttol: yes
(09:15:41) dr ttol: how come b.i2hub.com is never used? (or is it and we just dont know)
(09:16:12) David Gucwa: it should be used some of the time
(09:16:17) David Gucwa: I added it to the proxy list
(09:16:43) dr ttol: ok
(09:17:01) dr ttol: try importing via www.connectu.com/importer
(09:17:02) dr ttol: doesnt work
(09:17:16) David Gucwa: right
(09:17:24) dr ttol: do you think facebook changed their layout
(09:17:37) dr ttol: and are now catching IPs that are trying to do old layout
(09:17:39) dr ttol: and blocking them
(09:18:08) David Gucwa: I think the problem is the proxies
(09:18:57) David Gucwa: I'm trying to load connectu.com through the proxy fetch script and that won't load either
(09:19:11) David Gucwa: possibly we sent them so much traffic that they blocked us
(09:19:48) dr ttol: maybe
(09:20:04) dr ttol: add a.i2hub.com:8888 to the list
(09:20:17) dr ttol: but your proxychecker
(09:20:19) dr ttol: should catch that right?
(09:20:59) David Gucwa: it should
(09:21:05) dr ttol: alright, switch of tactic
(09:21:48) dr ttol: take a look at
(09:21:54) dr ttol: http://www.i2hub.com/checkproxies.com/proxycheck.php
(09:21:56) dr ttol: source:
(09:21:57) dr ttol: http://www.i2hub.com/checkproxies.com/proxycheck.phps
(09:22:20) dr ttol: use that proxychecker instead
(09:24:21) dr ttol: (modify that code to do what the old one was suppose to do)
(09:28:00) dr ttol: another issue: why does importer/facebook on status say green? i assume thats why you didnt get beeped
(09:28:10) dr ttol: (and CW and WW didn't see it)
(09:28:23) David Gucwa: that's a good question
(09:35:01) dr ttol: status?
(09:36:12) David Gucwa: comparing output of your checker to output of my checker
(09:36:40) dr ttol: ok, do we know if your facebook parser needs to be updated?
(09:36:46) dr ttol: or if this is solely a proxy issue
(09:37:08) David Gucwa: I don't know about the facebook format yet
(09:37:12) dr ttol: ok
(09:37:18) David Gucwa: I suspect that isn't the problem
(09:38:27) dr ttol: ok
(09:51:36) David Gucwa: hm

(09:51:39) dr ttol: hm?
(09:51:52) David Gucwa: I think they're doing a browser test or something
(09:51:59) David Gucwa: I can't seem to get to their page at all via links
(09:52:02) David Gucwa: proxies or not
(09:52:17) dr ttol: smart
(09:52:38) dr ttol: can we spoof
(09:52:52) David Gucwa: probably
(09:53:06) dr ttol: whats the most common user-agent
(09:53:51) David Gucwa: IE probably
(09:54:11) dr ttol: ok
(09:54:21) dr ttol: lets rotate the user-agents
(09:54:28) dr ttol: different ie versions
(09:58:18) dr ttol: lets do IE 6.0
(09:58:20) dr ttol: Firefox 1.0
(09:58:38) dr ttol: Safari 1.2
(10:00:20) dr ttol: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1)
(10:00:32) dr ttol: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)
(10:00:38) dr ttol: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1)
(10:00:45) dr ttol: Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.7.5) Gecko/20041107 Firefox/1.0
(10:00:52) dr ttol: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.0)
(10:00:59) dr ttol: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; .NET CLR 1.1.4322)
(10:01:05) dr ttol: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; .NET CLR 1.1.4322)
(10:01:23) dr ttol: Mozilla/5.0 (Macintosh; U; PPC Mac OS X; en) AppleWebKit/125.5.6 (KHTML, like Gecko) Safari/125.12
(10:01:35) dr ttol: Mozilla/5.0 (Windows; U; Windows NT 5.0; en-US; rv:1.7.5) Gecko/20041107 Firefox/1.0
(10:02:03) dr ttol: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.0.3705; .NET CLR 1.1.4322)
(10:02:10) dr ttol: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.0.3705)
(10:03:35) dr ttol: what's up?
(10:03:50) David Gucwa: trying to change it
(10:04:51) dr ttol: k
(10:04:59) dr ttol: so facebook knows
(10:08:55) David Gucwa: most likely
(10:09:02) David Gucwa: they might not know it's us
(10:09:07) dr ttol: ok
(10:09:17) dr ttol: lets spoof browsers
(10:13:35) David Gucwa: ok the spoof worked
(10:13:43) David Gucwa: i'm still having trouble grabbing profiles though
(10:14:24) dr ttol: did they change it
(10:14:34) David Gucwa: maybe
(10:22:39) David Gucwa: wow
(10:22:47) David Gucwa: they changed around their email display by quite a bit
(10:23:16) dr ttol: ?>
(10:23:18) dr ttol: ?
(10:23:21) dr ttol: explain
(10:23:23) David Gucwa: haha you have to look at it
(10:23:27) David Gucwa: if you log into the facebook
(10:23:35) David Gucwa: and look at the source code for your profile page
(10:23:49) David Gucwa: search the source for "Email:"
(10:24:00) dr ttol: haha
(10:24:03) dr ttol: javascript